



DATA MINING TECHNIQUES TO CLUSTERING CYBER CRIME DATA

Dr. Neelam Sahu¹ | Sagar Darokar¹

¹ Dept of Information Technology, Dr. C.V. Raman University Kota, Bilaspur (C.G.), India.

ABSTRACT

In data mining, Cyber Crime management is an interesting application where it plays an important role in handling of crime data. Cyber Crime investigation has very significant role of police system in any country. There had been an enormous increase in the crime in recent years. With rapid popularity of the internet, crime information maintained in web is becoming increasingly rampant. In this paper the data mining techniques are used to analyze the web data. This paper presents detailed study on classification and clustering. Classification is the process of classifying the crime type Clustering is the process of combining data object into groups. The construct of scenario is to extract the attributes and relations in the web page and reconstruct the scenario for crime mining.

KEY WORDS: Crime data analysis, classification, clustering.

I. INTRODUCTION:

Cyber Crime is one of the dangerous factors for any country. Cyber Crime analysis is the activity in which analysis is done on crime activities. Today criminals have maximum use of all modern technologies and hi-tech methods in committing crimes. The law enforcers have to effectively meet out challenges of crime control and maintenance of public order. One challenge to law enforcement and intelligence agencies is the difficulty of analyzing large volumes of data involved in criminal and terrorist activities. Hence, creation of data base for crimes and criminals is needed. Data mining holds the promise of making it easy, convenient and practical to explore very large databases for organizations and users. Developing a good crime analysis tool to identify Cyber crime patterns quickly and efficiently for future crime pattern detection is challenging field for researchers. A web page involving a crime can be thought of as a chain of actions with series of background attributes. We can analyze web information from the perspective of events and apply some research results related to the events to solve the problem of web crime mining.

II. RELATED WORK:

It is important to study the previous related works to both learn from the experience of others and to add something to our existing body of knowledge. Existing literature has been reviewed in three different areas: crime data mining, data extraction and data focus

Cyber Crime Data Mining: Data mining is defined as the discovery of interesting structure in data, where structure designates patterns, statistical or predictive models of the data, and relationships among parts of the data. The data mining techniques is using for some results on crime mining. This technique is applied to study Cyber crime cases, which mainly concerned entity extraction, pattern clustering, classification and social network analysis. This method used to get the data of criminals by using frequency occurrence of incidents.

Data Extraction Event: In web page extraction is the process to extract attributes and relationship. The idea of this event extraction is the method of retrieving the information from database. The proposed a method to append events for the concept of data mining techniques. Entity extraction has been used to automatically identify person, address, cases, and personal properties from police reports to the judge.

Data Focus: This research paper is on web mining of content, using clustering techniques the web mining focus on the text. In data focus clustering will convert nonlinear statistical relationship between high dimensional data into simple geometrical relationship in low dimensional display.

III. METHODOLOGY:

In this section, discuss about the methodology for the research.

Data Collection: The data set is the collection of field in the data from web pages on the internet. The data set which consists of the text from web pages and the pictures, videos or sound format will be ignored.

Preprocessing: A data preprocessing is a process that consists of data cleaning, data integration and data transformation which is usually processed by a computer program. It intends to reduce some noises, incomplete and inconsistent data. The results from preprocessing step can be later proceeding by data mining algorithm.

Clustering: Then the clustering techniques are used to fetch the information of criminals.

IV. OVERVIEW OF CYBER CRIME DATA MINING

Data Mining: Data mining deals with the discovery of unexpected patterns and new rules that are "hidden" in large databases. The use of data mining in this paper is to give the structured data from unstructured data of judge. In this paper the Data Mining techniques of crime in two directions they are

1. Classification of Cyber Crime
2. Clustering Technique of Cyber Crime

1. Classification of Cyber Crime

Cyber Crime: Crime is defined as "an act or the commission of an act that is forbidden, or the omission of a duty that is commanded by a public law and that makes the offender liable to punishment by that law". Crime is referred to as a comprehensive concept that is defined in both legal and non-legal sense.

Classification of Cyber Crime

- IPC ACT
 1. Offences by Public Servant
 2. Destruction of Electronic Record
 3. Cheating
 4. Forgery
 5. Data Theft
 6. Criminal Breach of fraud+Creadit, Debit Card
 7. Counterfeiting (Currency, Stamps, property)
 8. False Electronic Evidence
- IPC ACT
 9. Copyright Act 1957
 10. Trade Marks Act 1999

2. Clustering Technique of Cyber Crime

Clustering: Data clustering is a process of putting similar data into groups. A clustering algorithm partitions a data set into several groups such that the similarity within a group is larger than among groups. Clustering can also be considered the most important unsupervised learning technique; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. There are so many techniques used in clustering, in this paper only K-means algorithm is used.

K-Means Clustering Algorithm: K-means algorithm mainly used to partition the clusters based on their means. Initially number of objects are grouped and specified as K clusters. The algorithm clusters observations into K groups, where K is provided as an input parameter. It then assigns each observation to clusters based upon the observation proximity to the mean of the cluster. The cluster's mean is then recomputed and the process begins again. In this paper the use of K-means algorithm is the process of getting a structured data from a unstructured data. The working of algorithm is explained as follows:

- k : pre-determined number of clusters
- Algorithm (Step 0: determine value of k)

Step 1: Randomly generate k random points as initial cluster centers

Step 2: Assign each point to the nearest cluster center

Step 3: Re-compute the new cluster centers

Repetition step: Repeat steps 3 and 4 until some convergence criterion is met (usually that the assignment of points to clusters becomes stable).

V. RESULT:

The result of this crime management system is given below: Judgement of a Crime

Table 1: Unstructured data

case no 3475 XYZ is sentenced to jail for 24 years in the Durg for the Cheating crime of a person using Computer

Table 2: Structured data

Case no	3475
Name	XYZ
Crime Type	Cheating
Act	66D
Judgment	24 years
Location	Durg

The use of K-means clustering algorithm in this paper is to get the structured data from unstructured data from the data base. The prediction of this retrieving information from database is clear and understands to the user of the system. Predict the name, case, judgment, etc. from the judge. The judge gives the unstructured information of a criminal as judgment. The prisoner admin get the data of particular criminal information in a structured format.

VI. CONCLUSION:

Cyber Crime data is a sensitive domain where effective data mining techniques play a vital role for Cyber Crime analysis. In this paper the classify and cluster techniques are used to analyses the crime data from database. This technique is faster to get the data through web; the effective web mining is to get the unstructured data to structured data. The classifications of crime type are Cheating, Forgery, etc. and clustering of crime using k-means to retrieve the data.

REFERENCES:

1. D. E. Brown, "The regional crime analysis program (RECAP 1998) : A Frame work for mining data to catch criminals," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol.3, pp.2848-2853
2. IEEE proceedings Rushinek, A, Rushinek, SF (1993). "Using Experts for Detecting and Litigating Computer Crime". Managerial Auditing Journal. 8.7:19-22. Security Focus. Florida.
3. Anshu Sharma, Shilpa Sharma, An Intelligent Analysis of web Crime Data Using Data Mining, Internation Journal of Engineering and Innovative Technology, 2012.
4. S. Yamuna, N. Sudha Bhuvaneswari, Data Mining Technique to Analyse and Predict Crime, The International Journal of Engineering And Science, 2012.
5. Malathi. A, Dr. S. Santhosh Baboo, Anbarasi. A, An Intelligent Analysis of a City Crime Data Using Data Mining, Internation Conference on Information and Electronics Engineering, 2011
6. Devesh Bajpai, Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview, Journal of Environmental Science, Computer Science and Engineering & Technology, 2012
7. R.G Uthra Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview, Journal of Data Mining Technique to Analyze Crime Data, International Journal for Technological Research in Engineering , 2013